DOCUMENT RESUME

ED 098 251 TM 004 014

AUTHOR Bailey, John E., III
TITLE Predictive Ability of a Multi-Factor Grading System.
PUB DATE Jul 74
NOTE 17p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Academic Achievement; *Achievement Tests; Elementary School Students; *Grading; Mathematics Education; *Predictive Measurement; Scores
IDENTIFIERS *Multi Factor Grading

ABSTRACT

This paper investigates the relationship between a multifactor grading system and standardized achievement test scores. The study attempts to measure not only achievement but also motivation and rate of progress. Two hypotheses are tested: (1) Teacher measures of application, improvement, and grade level do not bear a significant relationship to standardized test scores in mathematics. (2) Teacher measures of application and improvement do not add, significantly, in prediction to that provided by the grade level and class. The procedure of the study is described in light of the definition of terms, sample, instrument, design, and data analysis. It is clear from the data that the grades at this private day school bear a strong relationship to achievement test scores and are a good predictor of them. However, neither application nor improvement, as measured, added significantly to the straight achievement measure. The reasons for, and implications of, this result are explained. (SE)

PREDICTIVE ABILITY OF A MULTI-FACTOR GRADING SYSTEM.

John E. Bailey III

Prepared For:

Dr. James Montgomery
North Broward Day School

1 July 1974
NOVA UNIVERSITY
Fort Lauderdale, Florida

The purpose of this study is to investigate the relationship between a multi-factor grading system and standardized achievement test scores. In recent years numerous segments of the population have attacked the traditional A-F grading system employed by the schools (Zimmerman, 1970; Miller, 1967; Glasser, 1969; and Milton, 1972). The criticisms of the traditional system are numerous, but one of the central criticisms has been that grades attempt to measure with one score several important and partially independent dimensions of behavior (Driscoll, 1972). There are a number of possible responses to the problems that the critics of the traditional system have surfaced. The most typical response has been to reduce the distinctions that are attempted, as in the various pass-fail systems. Other alternatives of this type have ranged from a computerized prose evaluations system (Giannangelo, 1974) to a system of only recording completes as students master behaviorial objectives (Zimmerman, 1970). A conceptual alternative to this strategy of making fewer distinctions is to attempt to measure the other important dimensions of the student in school. This alternative has been attempted by a private, day school. It attempts to measure not only achievement, but motivation and rate of progress as well.

RELATED LITERATURE

The study of the validity of school grades has had mixed results. A number of studies have shown that the high school grade point average is an excellent predictor of college

grades (Wilson, 1970; Farver, 1973; McCausland, 1974; and Wilson, 1971). These studies have shown this to be true for a variety of populations. By contrast Hoyt (1966) in a review of the literature dating back to 1917 has found that college grades bear little or no relationship to adult success as measured by a number of variables. Jackson and Lahaderne (1967) found only a small relationship between teacher awarded grades and standardized test scores in the sixth grade. Of course, all the above outlined studies were based on grades as defined by the traditional A-F system and thus do not attempt to measure separately motivation or rate of improvement.

HYPOTHESES

This study was designed to investigate two hypotheses. The null hypotheses are:

I. Teacher measures of application, improvement, and grade level do not bear a significant relation to standardized test scores in mathematics.

If hypothesis I is rejected, it is legitimate to test the hypothesis that

II. Teacher measures of application and improvement do not add significantly in prediction to that provided by the grade level and class.

Other questions of interest include:

1. Of the available teacher measures which set provides the best prediction combined with the fewest predictors?

2. Is there any difference in prediction between the prediction of the total math scores and subscores for

reasoning and computation?

PROCEDURE

Definition of Terms. Appendix A contains an example of the grade card that is used in this system. It is comprised of four components, application, improvement, grade level, and conduct. Application is graded on a five point scale along a dimension from seeks independent work to will not work. Improvement is a four point scale from accelerated to none. Grade level is a three point scale from above grade level to below grade level. Conduct is a satisfactory-unsatisfactory dichotomy.

Sample. The sample consisted of 31 white, middle socio-economic status children from a small, private day school in northern Broward County, Florida. Students were from 4th, 5th and 6th grades. Classes were small with 9, 8, and 14 students in each class respectively. The school emphasizes a modern, concept oriented approach to the study of mathematics. Sexes were approximately balanced.

Instrument. The standardized test used was the California Achievement Test, 1957 edition. It is widely used and has been favorably reviewed as to reliability and validity (Neidt, 1957). It consists of 11 scores, including mathematics reasoning, fundamentals and total, the scales used in this study.

Design. The school operates on a nine month school year divided into four 9 week quarters. The grades used were from the third quarter which ended in the last week of March 1973. Grades were awarded normally by the teachers.

Since this study was not even contemplated at that time, the subjects could not have been influenced by its existence. The achievement test was administered in the second week of April 1973 under the supervision of a Ph.D. psychologist and scored by the distributor. Grades from the quarter preceeding the administration of the achievement test were used to predict its results.

Data Analysis. The data was analyzed via the R01 Regression Analysis program using both generation and transformation of variables. This is the 1 September 1969 version compiled by George Peabody Computer Center.

Since it was not obvious that the various teacher measures can be assumed to be linear, all could be tested for linearity using the appropriate full and restricted models. Due to the limited range of the variables only grade level and class were so tested. Based on the results of that test, all other variables were assumed to be linear. Linear interactions were tested for grade level, improvement, and application with class.[7] Various alternative models were tested to determine which model was most parsimonious without being significantly poorer in prediction. Due to the small sample size, no attempt was made to determine if moderator groups existed.

RESULTS

For Hypothesis I, as can be seen from Table 1, all tested models were significantly better predictors than chance accounting for from 62% to 42% of the variance. Thus

Hypothesis I is rejected and it is appropriate to test Hypotheis II.

INSERT TABLE 1 ABOUT HERE

In terms of Hypothesis II, neither application or improvement scores, or their interaction with class add significant prediction to that provided by class and grade level. These have F-ratios of .102 and 2.415 and resulting in probabilities of .75 and .13 respectively. These results are contained in Table 2.

INSERT TABLE 2 ABOUT HERE

Table 3 contains the relative contribution by each of the predictors to the predict ı equation. The contributions by each predictor coincide with their use in the obtained prediction eq tions except for the non-inclusion of the interaction term for improvement and class. This non-inclusion can be explained by the very high correlation between this terms and the improvement score (r = .94). Its contribution is entirely included within the improvement score and thus it adds nothing to prediction. The prediction from the various models plus the number of predictors used is contained in Table 1.

INSERT TABLE 3 ABOUT HERE

In examining the question of whether there was any difference between the prediction of total score and concept and computation scores, important differences were found. Table 4 illustrates that all but two of the correlations between the predictors and the concept score are lower than those for total score and computation scores. Prediction

INSERT TABLE 4 ABOUT HERE

was also lower, ranging from a r^2 of 37% to 23%.

DISCUSSION

It is clear from the data presented here that the grades at this school bea . strong relationship to achievement test scores and are good predictors of them. Within the system employed, the teachers' perceptions of academic achievement in terms of total math scores are probably close to exhausting the predictive ability of the system.

However, in this sample, neither application nor improvement as measured, added significantly to the straight achievement measure. This is not to say that they are not useful or valid measures. Neither are designed to measure math achievement but are designed to measure the students' independence and improvement. To the extent that these are reflected in higher achievement and this higher achievement is accurately measured by the grade level score, neither would be expected to add to prediction. Whether they are useful measures must thus stand on their reliability and validity in terms of their original intention. The extent to which they are successful in this regard cannot be determined by this study.

However, for improvement there is some information that tempts one to believe that it is not so much a measure of improvement as of achievement. It is reported to parents in the form of the letters A-D. This alone would be a powerful suggestion to revert to the traditional system in so far as this particular measure is concerned. The fact that

is the single best predictor of academic achievement reinforces this idea.

Attempting to resolve the question of which predictor model is best presents a number of difficult tradeoffs. It is probably unnecessary to resolve this issue, other than by saying that some combination of these variables, treated as being continuous, will account for between 40-50% of the variance. For the purpose of reporting their children's achievement to parents, any of these relationships are strong enough. Equally if one wishes to research and does not have the relevant standardized tests, this data would suggest that grades are stongly enough related to be a practical alternative.

In comparing the traditional one grade model to this model which adds improvement and application scores, the outcome is complicated. Statistically no significant differences were found. However the difference in variance accounted for is fairly large (6% of the variance). Again since fine discriminations are not typically made on the basis of these scores, either model is acceptable. If any important decisions are being made, then it would probably be wise to use the fullest possible model. To the extent these additional scores add information about areas other than academic achievement, they may well be useful and important and justifiable on that basis.

In terms of the school's stated emphasis on the learning of mathematical concepts, a discrepancy was found. The measurement criteria are weighted far more strongly toward computational

skills than to reasoning or conceptual skills. In terms of achievement test scores, the students are in fact doing better in math reasoning than in computation, so the teaching emphasis may be there. But this success is not being reflected as accurately in the various grades as is computational ability.

Turning to the wider questions on which this study has bearing, there are two. One is the relation between teacher measures and student achievement. Clearly this study supports the idea that teacher measures are very clearly related to student scores on standardized tests in this school. Of course, standardized tests are not the same thing as student achievement. Students inevitably get a lot more, good and bad, out of the classroom than the specific subjects that they are taught. But there is strong evidence that standardized tests do in fact measure fairly well the degree to which the student has learned the skills and subjects that are explicitly being taught. Thus this study supports the idea that teacher measures do bear a strong relationship to student achievement in the formal subjects in the curriculum.

This brings up the second question. How does one explain the low or non-existent relationships between grades and measures of non-academic adult achievement. One possibility is to say that grades don't really measure what has been taught. This study for the reasons mentioned above does not support this contention. Grades are reasonably

good measures of what is in the curriculum. This leaves open the question of what then the schools do contribute. This study can add nothing to answering that question except to suggest that indicting the grading system is not the answer. Rather it suggests that the answer must be found by examining what is taught and in what way it should and does contribute to a successful life after school, however that may be defined.

SUGGESTIONS FOR FURTHER RESEARCH

The most direct continuation of this research would be the investigation of the relationships between grades and standardized test scores in other subjects. For each of these subjects the predictive ability of each model can be tested. If possible the study should use a larger sample which would allow for the testing of moderator variables.

Beyond the immediate question of the degree to which teacher measures are related to academic achievement, the far more important question is the degree to which academic achievement, however measured, is related to later life. Given the enormous effort in both time and money that is devoted to the schools and to research and development activities associated with them, knowing what the schools and the various subjects within the curriculum contribute to later life seems of critical importance. Multivariate prediction studies using either grades or standardized test scores to predict non-academic success could isolate the effect of each component of the curriculum.

BEST COPY AVAILABLE

BIBLIOGRAPHY

Driscoll, J.G. "Grade Beef, Not Students" The National Observer, 12:6, July 1, 1972.

Farver, A.S.; et. al. Longitudinal Predictions of Black and White University Student Grades. College Park: Maryland University, 1973.

Giannangela, D.M. & Lee, K.Y. "At Last: Meaningful Report Cards" Phi Delta Kappan, 1974, 55, 630-631.

Glasser, W. Schools Without Failure. New York: Harper & Row, 1969.

Hoyt, D.P. "College Grades and Adult Accomplishment." The Educational Record, Winter 1966, 70-75.

Jackson, P.W. & Lahaderne, H.M. "Scholastic Success and Attitude Toward School in a Population of 6th Graders" Journal of Educational Psychology, 1967, 58, 15-18.

McCausland, D.I. & Stewart, N.E. "Academic Aptitude, Study Skills and Attitudes and College GPA" Journal Of Educational Research, 1974, 67, 354-357.

Miller, S. Measure, Number and Weight: A Polemical Statement of the College Grading Problem. Knoxville: Learning Resources Center, University of Tennessee, 1967.

Milton, O. Alternatives to the Traditional. San Francisco: Jossey-Bass, 1972.

Moulds, H. "To Grade or not to Grade." Intellect, 1974, 102, 501-504.

Neidt, C.O. "California Achievement Tests, 1957 Edition." in Burros, O.K. The Fifth Mental Measurements Yearbook. Highland Park, N.J.: The Gryphon Press, 1959.

Wilson, K.M. The High School Average. Princeton: College Research Center, 1970. ED 081-812.

Wilson, K.M. The Utility of a Standard Composite for Predicting Freshman Grade Average in Eight Liberal Arts Colleges. Princeton: College Research Center, 1971. ED 081-807.

Zimmerman, R.R. & Halbert, T.D. Individualization of Instruction: Evaluation - A Roadblock. Education Monograph, 1970, No. 5.

TABLE 1

Variance Accounted For

Model	r^2	Number of Predictors
Grade and Class Non-Linear	.6157*	12
Full Linear	.5380*	8
Improvement-Interaction	.5141**	6
Improvement w/o Interaction	.5127**	5
Application-Interation	.4838**	6
Application w/o Interaction	.4695**	5
Class and Grade-Linear Interaction	.4674***	4
Class and Grade-Linear w/o Int	.4254***	3

*p .05 **p .01 ***p .001

TABLE 2

Significance Test

Full Model	Difference in r^2	F Ratio	Probability
Improvement-Interaction	.0467	1.202	.3175
Improvement w/o Interaction	.0453	2.415	.1288
Application-Interaction	.0164	.397	.6818
Application w/o Interaction	.0021	.102	.7502

Note: All models tested against Class and Grade-Linear Interation.

TABLE 3

Contribution Coefficient

Predictor	Contribution
Application	.42
Improvement	.78
Grade Level	.58
Conduct	.12
Class	.70
Application-Class Interaction	.05
Improvement-Class Interaction	.48
Grade Level-Class Interaction	.32

TABLE 4

Test - Predictor Correlations

Predictor	Concept	Computation	Total
Application	.13	.24	.22
Improvement	.39	.43	.42
Grade Level	.27	.35	.32
Conduct	.01	.16	.06
Class	.24	.36	.38
Application-Class Interaction	.04	.04	.03
Improvement-Class Interaction	.27	.27	.26
Grade Level-Class Interaction	.17	.21	.17

APPENDIX A

Multi-Factor Grade Card

KEY TO GRADING

Application:

1. He/She seeks out work independently.
2. He/She works independently if assigned.
3. He/She works with moderate supervision.
4. His/Her work requires constant supervision.
5. He/She will not work.

Improvement:

- A Accelerated
- B Sufficient
- C Insufficient
- D None

Grade Level:

- + Above Grade Level
- √ At Grade Level
- — Below Grade Level

Conduct:

- S Satisfactory
- U Unsatisfactory

	First Quarter				Second Quarter				Third Quarter				Fourth Quarter			
	Application	Improvement	Grade Level	Conduct	Application	Improvement	Grade Level	Conduct	Application	Improvement	Grade Level	Conduct	Application	Improvement	Grade Level	Conduct
SOCIAL STUDIES																
MATH																
SCIENCE																
SPANISH																
Reading																
Writing																
Comprehension																
Oral																
ENGLISH GRAMMAR																
Word Structure																
Composition																
Oral Expression																
READING																
Comprehension																
Phonetics Phraseology																
SPELLING																
No. of Days Present (out of 45).																
No. of Days on Time																

The following activities are provided on a less formal basis. The criteria for an evaluative judgment are participation, attitude and skills. Since these would be subjective judgments with few objective bases, they shall remain un-graded.

PHYSICAL EDUCATION ART MUSIC